The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Signed

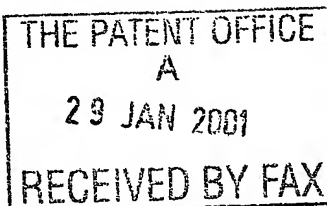Dated   3 0 JAN 2002

THIS PAGE BLANK (USPTO)

Pate)Form 1/77

Patents Act 1977
(Rule 16)

Request for grant of a patent

*(See the notes on the back of this form. You can also get an
explanatory leaflet from the Patent Office to help you fill in
this form)*

THE PATENT OFFICE
A
29 JAN 2001
RECEIVED BY FAX

The Patent Office

Cardiff Road
Newport
South Wales
NP10 8QQ

---

1. Your reference   **ℜFS | VOICEWEB**

29JAN01 E601609-1 D01463
P01/7700-0.00-0102230.0

2. Patent application number   **0102230.0**
   (The Patent Office will fill in this pa   **29 JAN 2001**

3. Full name, address and postcode of the or of
   each applicant *(underline all surnames)*

   Hewlett-Packard Company
   3000 Hanover Street
   Palo Alto
   CA 94304, USA

   Patents ADP number *(if you know it)*

   If the applicant is a corporate body, give the
   country/state of its incorporation

   Delaware, USA
   4965 8800l

4. Title of the invention   Sound related systems and methods

5. Name of your agent *(if you have one)*

   "Address for service" in the United Kingdom
   to which all correspondence should be sent
   (including the postcode)

   Robert F Squibbs
   Hewlett-Packard Ltd, IP Section
   Filton Road
   Stoke Gifford
   Bristol BS34 8QZ

   Patents ADP number *(if you know it)*   792818700l

	Country	Priority application number *(if you know it)*	Date of filing *(day / month / year)*
6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number			

	Number of earlier application		Date of filing *(day / month / year)*
7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application			

8. Is a statement of inventorship and of right
   to grant of a patent required in support of
   this request? *(Answer 'Yes' if:*
   a) *any applicant named in part 3 is not an inventor, or*
   b) *there is an inventor who is not named as an applicant, or*
   c) *any named applicant is a corporate body.*
   See note (d))

   Yes

Patents Form 1/77

Pate } Form 1/77

9. Enter the number of sheets for any of the
    following items you are filing with this form.
    Do not count copies of the same document

Continuation sheets of this form	
Description	20 +7 = 27
Claim(s)	—
Abstract	—
Drawing(s)	—

10. If you are also filing any of the following,
    state how many against each item.

    Priority documents

    Translations of priority documents

    Statement of inventorship and right
    to grant of a patent (Patents Form 7/77)

    Request for preliminary examination
    and search (Patents Form 9/77)

    Request for substantive examination
    (Patents Form 10/77)

    Any other documents
    (please specify)

11.   I/We request the grant of a patent on the basis of this application.

    Signature *R F Squibbs*
    Robert Francis Squibbs   Date 29 January 2001

12. Name and daytime telephone number of   K Nommeots-Nomm   Tel: 0117-312-9947
    person to contact in the United Kingdom

Warning
*After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication
or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You
will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the
United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting
written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the
United Kingdom for a patent for the same invention and either no direction prohibiting publication or
communication has been given, or any such direction has been revoked.*

Notes
a) *If you need help to fill in this form or you have any questions, please contact the Patent Office on 08459 500505.*

b) *Write your answers in capital letters using black ink or you may type them.*

c) *If there is not enough space for all the relevant details on any part of this form, please continue on a separate
    sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be
    attached to this form.*

d) *If you have answered 'Yes' Patents Form 7/77 will need to be filed.*

e) *Once you have filled in the form you must remember to sign and date it.*

f) *For details of the fee and ways to pay please contact the Patent Office.*

Patents Form 1/7

1

SOUND-RELATED SYSTEMS AND METHODS

The present invention relates to systems and methods of interacting with sound-based services.

Adaptive use of bandwidth in 3D audio sessions

Abstract - Techniques for reducing the network transmission bandwidth required to implement a 3D audio interface to remote services on a user device.

Background - A 3D audio interface uses spatialisation processing of sounds to present services in a synthetic but realistically plotted three dimensional audio field. Sounds, representing services and information could be placed at different distances to the front, rear, left, right, up and down. An example of a service might be a restaurant. A pointer to the restaurant (the equivalent of a hyperlink) is placed in the audio field. There are several ways to represent audio hyperlinks; repeating a service name (e.g. name of restaurant), perhaps with a short description of the service, an earcon for the service (e.g. a memorable jingle or noise), or perhaps an audio feed from the service.

Such an interface relies upon a high quality audio interface that is capable of rendering a 3D audio field. In applications where the user device is interacting with a remote 3D audio application via a relatively low bandwidth audio channel (e.g. a wireless telephony link) or indeed any channel in which lossy audio codecs are employed, it is likely that the channel will degrade the 3D nature of the audio, perhaps to the point of masking any perception of 3D positioning of sounds (figure 1).

This problem does not occur if the multiple audio signals that form components of the 3D audio interface are transmitted independently to the user device, where they are then combined (figure 2). The 3D audio is therefore not subjected to the lossy transmission channel. However, such a system potentially requires a larger total bandwidth to carry the multiple component audio signals. In many network applications, particularly mobile wireless networks, the bandwidth of the access channel is a limited and expensive commodity.

2

The problem to be solved is how to minimize the bandwidth required to transmit the component audio signals to the user device, whilst preserving the facility of a high quality 3D audio interface.

5 **Description:** We describe three techniques that may be used to address this problem.

Technique 1: In this technique the 3D audio processing is performed at the user device. We observe that at any point in time a user will have a primary focus (or foci) within the 3D audio interface. For example, the user may have selected the
10 aforementioned restaurant service and be interacting with it. This primary focus may be rendered at the position "straight ahead" in the 3D audio field. It is desirable that this primary focus be rendered as a relatively high quality audio signal. However, other services that are not currently a primary focus might be adequately represented in the 3D audio field by a lower quality audio signal. It is therefore possible to reduce
15 the bandwidth required for a component audio signal in the transmission channel by selecting a lower bitrate (which generally means lower quality) codec for that component while it is not the primary focus of the user (we note that as the quality of an audio signal is degraded, it is still possible for that audio signal to be placed accurately in a 3D audio field).
20

When a user (or some programmatic operation) selects a service as a primary focus, the corresponding audio signal is switched to use a higher bit-rate codec. At the same time it is envisaged that services ceasing to be a primary focus are switched to a lower bit-rate codec. In this way the total bit rate required to transmit all audio components
25 is reduced.

This technique might be implemented using variable bitrate codecs (which already exist) and a control channel to signal the required bitrate/quality from the user device to the source of each audio component (figure 3). Such signalling might also be
30 present in order to control codec bitrate for the purposes of network congestion control or adaption to channel conditions.

3

Technique 2: This technique is a variation on technique 1. The audio sources that are not currently the primary focus of the user send an audio sample to the user device (as opposed to a continuous audio stream). At the user device this audio sample is stored and then repeated in the audio mix at the appropriate 3D position. The bandwidth

5   required by these audio sources is thereby very small.

When a source becomes the primary focus of the user, it is requested to transmit a continuous stream of audio to the user device, and this stream replaces the repeating sample in the 3D audio field. As in technique 1, a feedback control channel is

10   necessary between the user device and the component audio sources.

The audio samples may be cached on the user device and re-used when a channel ceases to be the primary focus, analogous to a service being "minimized" as an icon on a visual desktop.

15

Technique 3: In this technique some of the 3D audio processing is performed at the source of the audio components (or at some network node that aggregates audio components). As discussed above, subsequent transmission across a lossy channel will result in a degradation of the 3D spatialisation of the audio interface.

20

To combat this degradation, a low bandwidth "tracer" for each audio component is transmitted to the user device in addition to the 3D audio signal. The tracer would comprise a description of the component's intended position in the 3D audio field and a low bitrate version of the audio component. The low bitrate audio component in the

25   tracer is of much lower quality than the main 3D audio signal and its components. However, due to its correlation with the original audio component, it is sufficient to allow association by the human ear with the corresponding component in the main 3D signal.

30   At the user device the tracers are used to add the low bitrate (low quality) versions of each component to the 3D audio field with high positional accuracy (we note that even poor quality audio signals may be positioned with high accuracy in a 3D audio field). The combination of a high quality signal with low 3D audio positional accuracy, and a set of low quality audio signals with high 3D audio positional

4

accuracy results in the restoration of the human perception of 3D position to the degraded 3D audio signal.

An advantage of this technique is that the 3D audio channel may be generated in a
5 network based device, thereby reducing the bandwidth required in the access network to that of a stereo channel. Those devices capable of rendering 3D audio may request the additional tracers whilst other devices simply render the main stereo channel. The bandwidth required to transmit the tracers is small compared to that required to transmit all component signals.
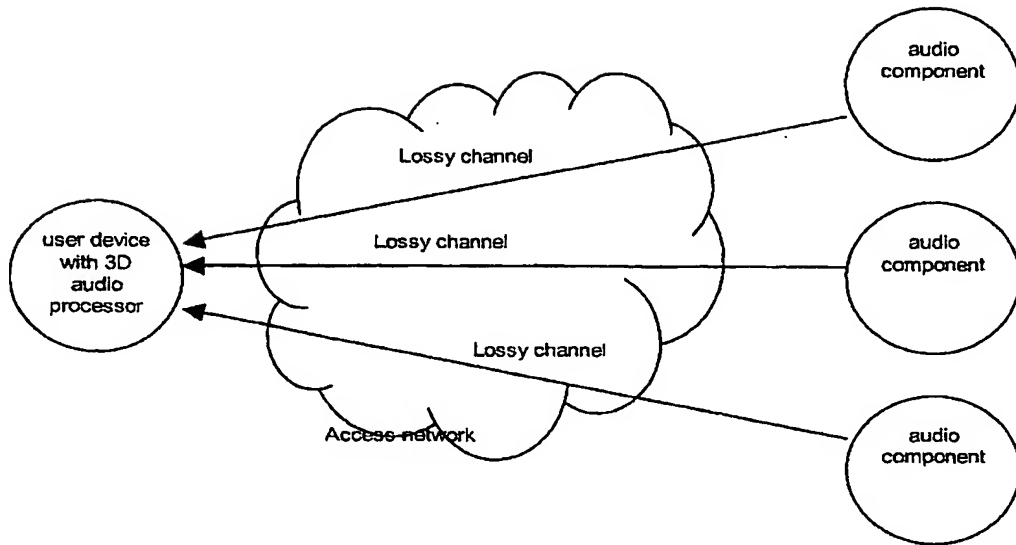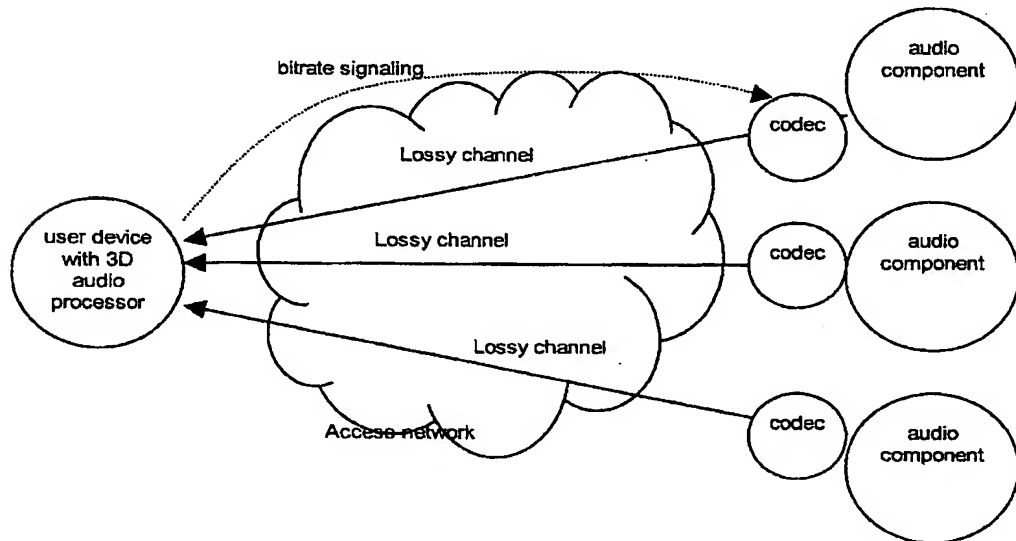
10
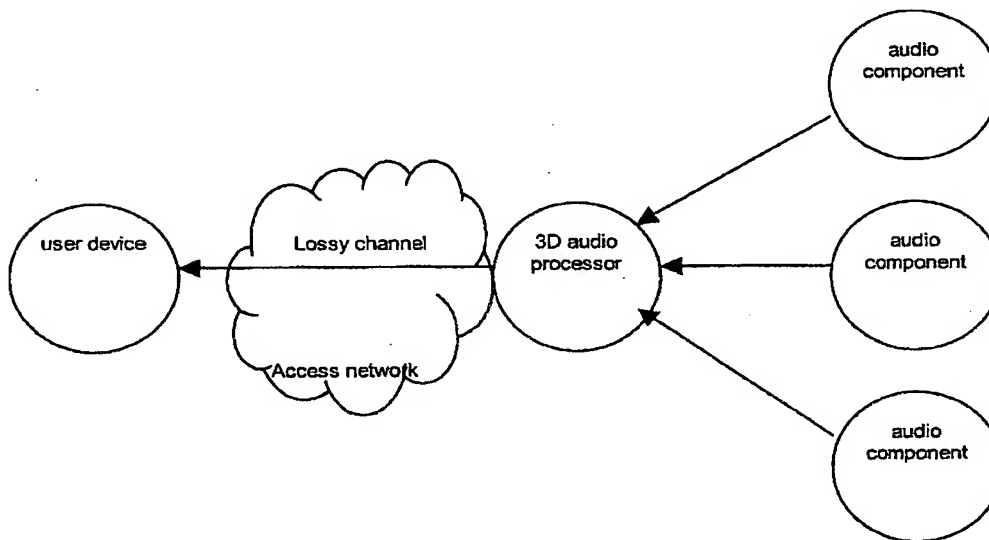
15   Figure 1: 3D processing performed remotely

5

Figure 2: 3D processing performed locally

5   **Figure 3: 3D processing performed locally with variable bitrate codecs**

6

3D audio cursor UI

Abstract: User interface technique to navigate and select audio objects placed in a 3D audio field.

5   **Background:** A 3D audio interface uses spatialisation processing of sounds to present services in a synthetic but realistically plotted three dimensional audio field. Sounds, representing services and information could be placed at different distances to the front, rear, left, right, up and down. An example of a service might be a restaurant. A pointer to the restaurant (the equivalent of a hyperlink) is placed in the

10   audio field. There are several ways to represent audio hyperlinks; repeating a service name (e.g. name of restaurant), perhaps with a short description of the service, an earcon for the service (e.g. a memorable jingle or noise), or perhaps an audio feed from the service. The problem is how to navigate and select services from the ones presented to the user.

15

Description: One technique to select services is to have a 3D audio cursor that the user can move in the audio field. The problem is how to feed back to the user the position of the cursor relative to the services to be selected. The 3D cursor could make a distinctive noise, such as a hum (a hum is relatively easy for a user to perceive a

20   location in a 3D audio field). The user can move the 3D cursor using a device such as track ball. As the cursor approaches service objects, the characteristics of the sound of the cursor would change; for example, the hum might get louder and higher pitched. In addition, the services near the cursor may get louder, change pitch, make an additional sound (a kind of sound aura), repeat a sound with an increasing frequency.

25   Only the services near the cursor would be audible, to minimize audio clutter. The details of the sound changes of the objects would depend on their relative position to the 3D cursor; for example objects to the right may increase in pitch, to the left may decrease. When the 3D cursor was at the selection point of a service, the service may make a noise, or say something such as "ooohhh". To activate the service at the

30   selection point, the user could simply press a button, or speak the unique service name. In fact the service name can be used at any time, the audio cursor makes it easier to figure out the set of services to select from.

7

Context Dependent Acoustic Properties

Abstract: Changing the acoustic properties of responses based on context.

Background: Speech is very good at cutting to the point, but tedious when used to
5   read a long list of options

Description: The idea is to vary the acoustic properties of a computer generated
response to a question, in accordance with the number of options returned. If there are
a large number of options, as returned from a web search, then the response would
10   sound as if it had come from a large room. If the number of responses is small, then
the acoustic properties of the response would reflect those of a single voice in a small
room. Using audio cues to gaining an impression of scale has the potential to
streamline a dialogue containing many or few options.

15

Example:

   User: "How many cinemas are there in Bristol?"

       If there is only one:
20   System: "Bristol only has one cinema" in a voice as if from a broom cupboard.

       If there are many:
   System: "There are 15 cinemas in Bristol" in a voice, as if from a cathedral.

25

8

Ad hoc 3D audio devices

Abstract: Ad hoc use of audio output devices to render audio and speech output from a network-based audio and speech server.

5   **Background:** As a mobile user moves from place to place in the environment, the user will encounter audio input and devices. For example, a car typically has a near hi-fi quality audio system with multiple speakers; a car of the future may have microphones built in. Similarly, a home has hi-fi surround sound audio equipment and may have several microphones throughout the house. Depending on the situation, it

10   could be useful for the user to be able to use these devices instead of using a mobile phone.

Description: The invention is to give audio devices an Internet connection such that they can connect to the users audio/speech server in the network. The audio device

15   would be able to receive a connection endpoint address from the user (1), in the form of a URL, via an infra-read or short-range radio transmitter; user authentication information is also passed across. This URL and authentication information is passed on to the Connection Manager of the device (2), which in turn, contacts the audio/speech server over the Internet to negotiate how to connect the audio stream

20   into the users session (3). The negotiation resolves the network technology to be used to carry the audio (e.g. a phone network or the Internet), and the details of the audio data to be exchanged which obviously depends on the device capabilities. This includes the number of audio channels, the codecs to be used, the allocated bandwidth, the number, type and position of speakers, 3D audio processing

25   capabilities, etc. For example, the negotiation may decide that all mixing and 3D spatialisation processing is to take place in the audio server and carried in a high-bandwidth stereo connection. Alternatively, that individual streams are sent to the device, together with 3D positioning information, and that the mixing and 3D spatialisation processing is done locally on the device. After negotiation, real time

30   audio data is exchanged between the devive and the server (4) over the network of choice. Other, non-real-time audio data may also be exchanged (5). The device may have additional specialised local processing capabilities that can be used to optimise bandwidth utilisation. For example a car may have its own TTS engine; the speech server could simply send textual strings and 3D position information to it over a low

9

bandwidth connection. Alternatively, if a file is to be played, the URL to the audio content could be passed to an audio player in the car, and mixed and spatialised locally.

5   The notion of negotiating where and how the 3D spatialisation processing takes place is important from the point of view of scalability and bandwidth utilisation. Also other device characteristics are important, such as on-board TTS or audio playback.

10

10
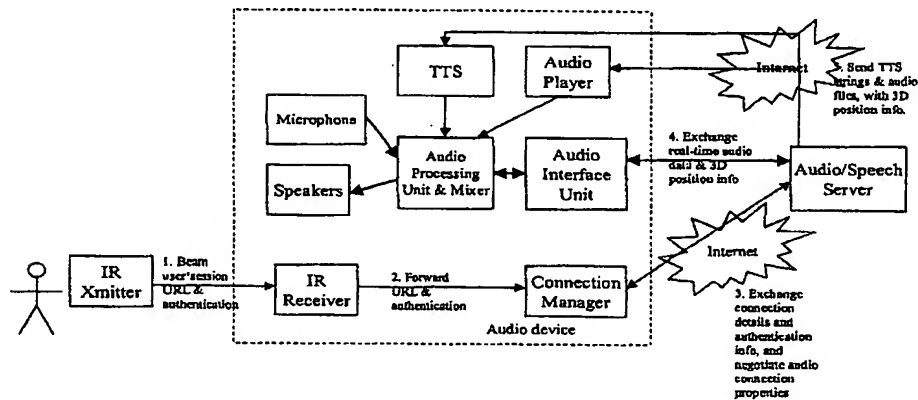
Minimize audio clutter in 3D audio interfaces

Abstract: User interface technique to minimize clutter in an audio user interface, in order to maximize the amount of information that can be presented to a user without confusion.

5

Background: A 3D audio interface uses spatialisation processing of sounds to present services in a synthetic but realistically plotted three dimensional audio field. Sounds, representing services and information could be placed at different distances to the front, rear, left, right, up and down. An example of a service might be a

10   restaurant. A pointer to the restaurant (the equivalent of a hyperlink) is placed in the audio field. There are several ways to represent audio hyperlinks; repeating a service name (e.g. name of restaurant), perhaps with a short description of the service, an earcon for the service (e.g. a memorable jingle or noise), or perhaps an audio feed from the service. The problem is how to maximize the number of services that can be

15   presented without overloading the audio perception and cognitive capabilities of the user. The problem applies both to audio augmentation of the real world, and to the general case of multiple audio services attempting to attract a users attention.

Description: The solution is to arrange for services to periodically fade in for a

20   small number of seconds and then fade away again. The audio platform manages this process such that only a small number of services are presented at any one time. The 3D position of objects is preserved and is consistent from one fade sequence to another. In an extreme case, only a single service sounds at any one time (analogous to TDM), but for many users a small number could sound simultaneously (analogous

25   to FDM). The user would build a mental model of the current set of services, and their location. The user could optionally refresh a portion of the audio field; for example the user might say "refresh 2 o'clock"; this would cause the selected set of services to repeat their name, earcon, or description.

30   All this helps to make 3D audio navigation and selection of services a practical possibility

11

Managing augmented reality in 3D audio interfaces

Abstract: User interface technique to manage the presentation of an audio augmentation of reality, integrated with other audio services that have no relation to the physical world.

5

Background: A 3D audio interface uses spatialisation processing of sounds to present services in a synthetic but realistically plotted three dimensional audio field. Sounds, representing services and information could be placed at different distances to the front, rear, left, right, up and down. An example of a service might be a restaurant. A

10   pointer to the restaurant (the equivalent of a hyperlink) is placed in the audio field. There are several ways to represent audio hyperlinks; repeating a service name (e.g. name of restaurant), perhaps with a short description of the service, an earcon for the service (e.g. a memorable jingle or noise), or perhaps an audio feed from the service.

15   As well as allowing augmentation of the real world, a 3D audio interface is also useful to manage services that have only logical, rather than physical relationships. A problem, therefore, is that the user does not always want to focus attention on the set of objects that represent services in the real world, and would like to make a clear distinction between logical and augmented services. Nevertheless, the user may like

20   to hear important notifications from services in the augmented space.

Description: The solution is to collapse the augmented space to a single point when the user does not want to focus attention on it, and give it a fixed logical position in the 3D audio field. All important sounds from the augmented space that pass a set of

25   filter criteria could be mixed and presented from the allocated logical position. If the user hears something that attracts his attention, the augmented space could be selected. This would cause the augmented space to expand to fill the whole audio field. Now all objects in the augmented would be layed out in their appropriate location corresponding to the physical relationships.

30

12

Efficient presentation of information in a 3D audio interface

Abstract: User interface techniques to manage, present and interact with services in a 3D audio field.

5   **Background:** A 3D audio interface uses spatialisation processing of sounds to present services in a synthetic but realistically plotted three dimensional audio field. Sounds, representing services and information could be placed at different distances to the front, rear, left, right, up and down. An example of a service might be a restaurant. A pointer to the restaurant (the equivalent of a hyperlink) is placed in the

10   audio field. There are several ways to represent audio hyperlinks; repeating a service name (e.g. name of restaurant), perhaps with a short description of the service, an earcon for the service (e.g. a memorable jingle or noise), or perhaps an audio feed from the service.

15   The problem to be solved is how to present information to a user via a 3D audio interface in an intuitive way, in order to maximize the set of services and applications that the user can deal without suffering cognitive overload.

Description: A number of user interface techniques are described, that together

20   reduce the cognitive load on the user to acceptable levels. User interface techniques are also described that deal with perceptual limitations in locating 3D positions and discriminating simultaneous sounds. Briefly the techniques include:

- Direction can be used to encode importance, for example in search results. E.g.

25   importance goes clockwise.

- Rotation of audio environment. User can spin all objects to bring sounds in a specified quadrant or position to the front, e.g. to select them, or to get better spatial resolution.

- User can mute segments of audio space.

30   - Position objects in logical layers in a cylinder. Objects are placed left-right around a circle on the horizontal. There are many circles, brought from layers up or down an imaginary cylinder, that each represent a different context.

13

- Position as a mnemonic. For search results or lists are layed out in 3d, to make selecting and remembering easier. For example, user says "select front left".

- Use position to represent work areas that can be selected or expanded. E.g. behind is unimportant background tasks, left-rear is e-mail, front-below is telephone calls, 5 front-up is interruptions such as reminders, etc. User learns to judge importance of sounds from direction.

- Use 3D positioning to fill a space with search results from a Web search. The relevance of the search is ordered in a clockwise direction, the received time or logical grouping of the responses is positioned up or down.

10 - Sounds can be processed to mean different things. E.g. sounds that leak from a room might echo, continuous background information might whisper, important information might have a lower pitch.

- Fill position slots for services in a clockwise manner, to follow the order in which the user starts services. Services maintain this fixed position to allow user to build 15 up a metal picture. When user selects a service to speak to, it moves to the front position, when switch to another service it returns to its original position.

14

Using different characters to announce and select services

Abstract: Using different characters (speakers) to announce and select services in an audio environment.

5   **Description:** In a workplace environment, people often occupy a fixed location (i.e. a cube). An employee knows his location relatively to other people. For instance, Joe may know that Mary is in the cube in front of me, Luke on his right, etc. Therefore, when Joe hear Mary's voice, Joe knows it is likely to come from the cube in front of him. By analogy, the user audio environment could be divided into portions (for

10   instance left, right, front, back). To each partition is associated a specific voice that is familiar to the user (i.e. Mary or Luke's voice). When a service appears in the environment, it notifies itself with a spoken message rendered in a special voice corresponding to the location of the service. For instance, a TTS (text-to-speech converter) with Luke's voice announces a new pizza delivery on Joe's right. The user

15   can then select a service by addressing a speaker and specifying the service he wants to execute. For instance, saying, "hey Luke, put me through the pizza service please."

This mechanism provides further clues concerning the location of the service in the

20   audio space. By recognising the speaker of the message, the user can locate the service in the audio space. It helps him to build a mental representation.

15

Distinguish real from virtual sounds

Abstract: Techniques to distinguish virtual sounds from real sounds in the environment.

5   **Background:** In the near future, the use of mobile devices to augment reality with additional information and services is likely to increase dramatically. Cooltown is one example of this. Presenting audio information to the user, perhaps in an 'always-on' 3D spatialised audio field, may become common place. As the audio capabilities of mobile devices improve, because of high bandwidth stereo connections, it becomes

10   increasingly difficult for the user to distinguish virtual sounds from sounds in the environment, resulting in confusion and annoyance.

    Description: The solution is to provide several mechanisms that together help the user form a clear distinction between the real and virtual. These techniques can be

15   divided into two broad groups. The first group involves processing the audio to make it sound unnatural. Briefly, these are to deliberately distort, muffle, add noise, frequency shift, bandwidth limit, add echo, change volume, or otherwise process the sound. The second group is a set of techniques to make the 3D spatialisation of the sounds different from the behaviour of objects in the real world. One example is to

20   explicitly to make the audio world fixed relative to the users head and movement (so as the user moves, the real-world objects move relative to the virtual). The techniques would be a) have objects in the virtual world only update their position at periodic intervals (e.g. every 30s), b) allow the user to explicitly freeze or update the position of objects, c) position objects relative to current position, not current orientation

25   (objects to North are in front, to south behind).

16

Represent Closeness Using Speaking Style and Vocabulary

Abstract: Represent closeness of service in an audio environment using modulation in speaking style and vocabulary.

5   **Description:** Services appearing in the audio environment notify the user with a synthesized message. The idea is that the content of the message and the way it is rendered depends on several parameters: the distance (in the audio space) between the user and the service, and how familiar the user is with this service.

10   A service located far from the user will try to shout to attract his attention, whereas a closer service would just whisper. Another idea is that services use a lexicon that depends on how far they are form the user. For instance, a close service could address the user using his first name. For instance "hey John, I am HP invent". The further the service is from the user, the more formal the presentation is. "Hello Mister Smith" is
15   used by a more distant service, "Excuse me sir" by even a more distant service, or even "Hello" for a very far service.

A similar mechanism is used to express how familiar a user is with a service. Services that are used all the time (i.e. e-mail, pizza booking) should present themselves with a
20   familiar voice or speech style. For instance: "hey buddy, do you fancy a pizza tonight?"

17

Use of Volume to Represent Closeness

Abstract: Using volume of audio output to represent closeness.

Background: As the number of voice based services increase, so to will the
5   congestion in the audio domain. This disclosure addresses this problem by
partitioning the audio space based simply on volume.

Description: As voice markup languages, 3rd generation mobile phone and location
based services become more pervasive, structuring the virtual and augmented audio
10   domain will assume greater importance. The idea behind this disclosure is simple:
services convey their physical or virtual proximity to a user through the loudness of
their voice.

15   **Example:**

Distance of Service from User

18
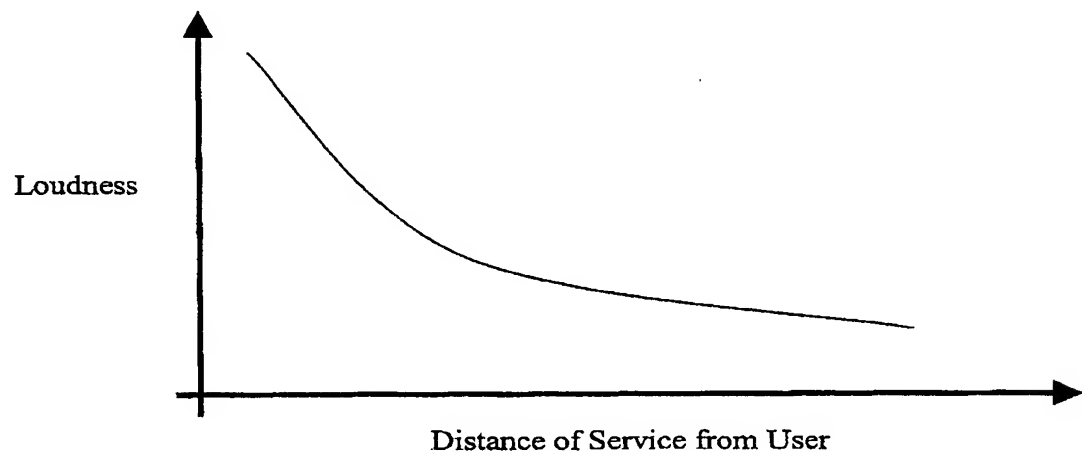
Use of Loudness to Address Service

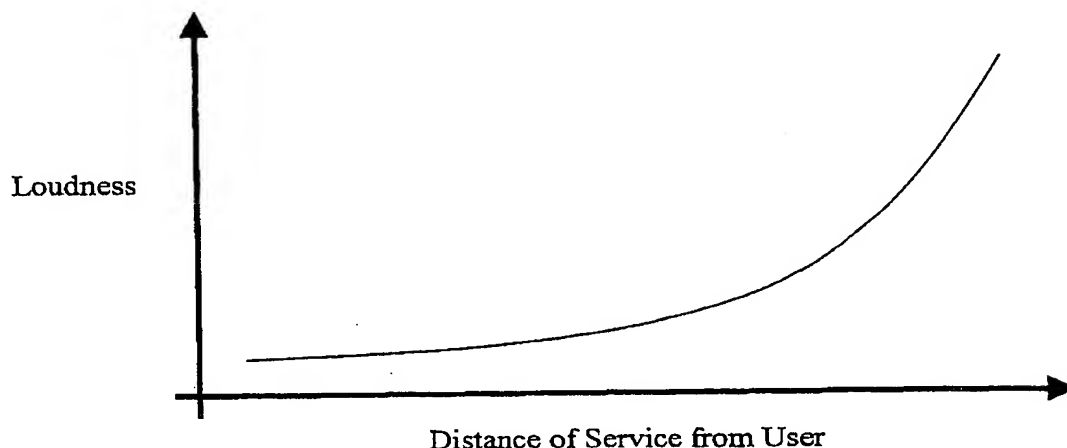**Abstract:** Using the loudness of a speakers voice to judge which service is being addressed.

5

Description: As voice markup languages, 3rd generation mobile phone and location based services become more pervasive, structuring the virtual and augmented audio domain will assume greater importance. The idea behind this disclosure is simple: when addressing a service, the user speaks louder to more distant services and quieter to closer services. The volume of the user's voice provides a very simple mechanism for judging distance.

10

In an environment where there are many services, it is important to reduce ambiguity and confusion during speech recognition by judging which service the user is addressing. The volume of the user's voice provides a very simple mechanism for judging distance. Services further away are addressed in a louder voice than closer services. The result is to reduce the search space for speech recognition, which will in turn increase performance.

15

20   **Example:**

Loudness (y-axis) vs Distance of Service from User (x-axis)

19

Use of Head Position to Address Service

Abstract: Using the head position to judge which service is being addressed.

Description: As voice markup languages, 3rd generation mobile phone and location

5   based services become more pervasive, structuring the virtual and augmented audio domain will assume greater importance. The idea behind this disclosure is simple: services are organized in 3D or 2D audio space around the user. If the user wishes to address a service on their right, they simply turn their head to the right and speak. The position of the head could be detected and resolved by a network of transmitters and

10   receivers on the user's head and shoulders. A Bluetooth headset with receives on the shoulder could be adapted to realize head position. Alternatively, microphones on each shoulder can be used to resolve the direction of speech.

Advantages: In an environment where there are many services, it is important to

15   reduce ambiguity and confusion during speech recognition by judging which service the user is addressing. On approach to this problem is to compare the position of a user's head, with the position of services in a 3D or 2D audio domain. The result is to reduce the search space for speech recognition, which will in turn increase performance.

20

Example: User wishes to speak to Service A on their left.

Service A   Service B

Direction
of Head

2D or 3D
audio space

20
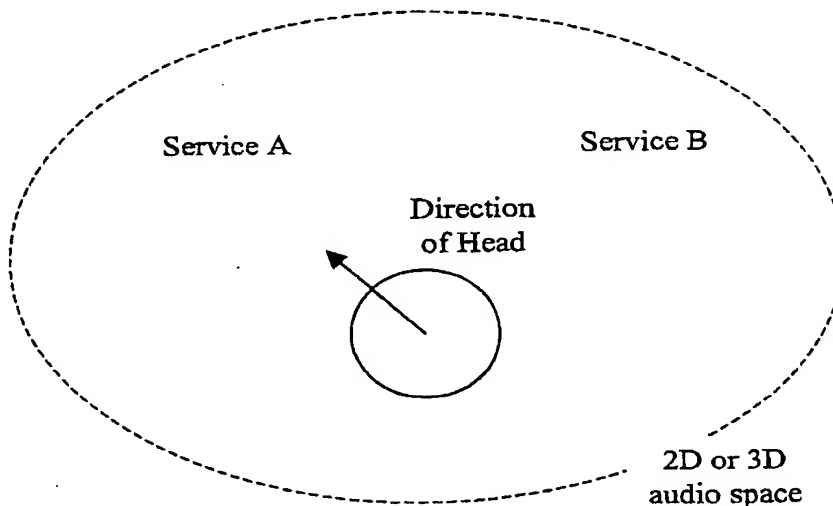
Use of Distance & Loudness to Indicate Relevance, Importance or Secrecy

**Abstract:**   Use of distance and/or loudness as an indication of relevance, importance or secrecy.

5   **Description:** As voice markup languages, 3rd generation mobile phone and location based services become more pervasive, structuring the virtual and augmented audio domain will assume greater importance. The idea behind this disclosure is simple: the importance or relevance of a service to the user is referenced in the perceived distance in the audio space around the user. Services of relevance, or secret in nature, would

10   appear closer to the user and may even whisper. This technique would work well in 1,2 or 3 dimensions depending on the capabilities of the device used to render the audio.

Background: As the number of voice based services increase, so to will the
15   congestion in the audio domain. This disclosure addresses this problem by partitioning the audio space in terms of the importance.

Example: Service A is very important or relevant to the user, and as such, sounds closer.

20

Service B
Not Important

Service A
Very Important

User's Head

1D, 2D or 3D
audio space

21
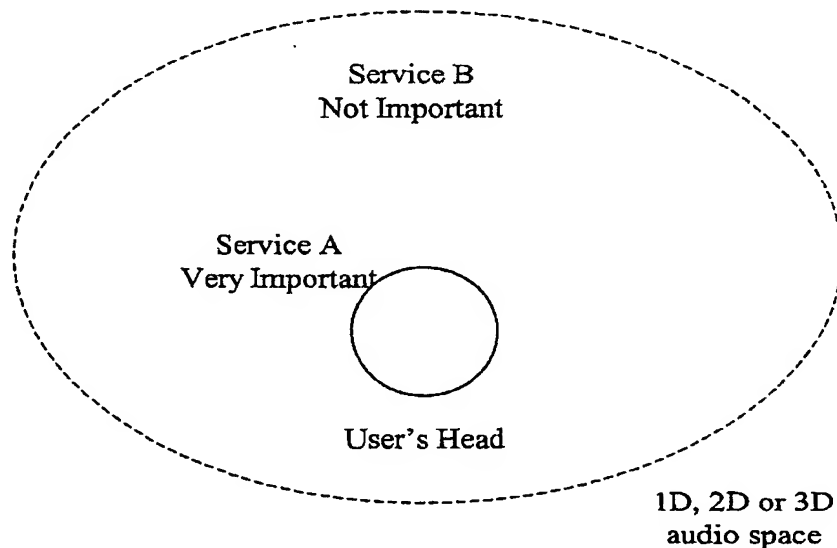
Further Features

(usable alone or in combination)

1. Use volume of audio output to represent closeness.
2. Use volume of user's speech output to help figure out which service is being addressed – the user speaks more loudly to distant objects. Associated idea is to use other mechanisms to code distance – suggestions were gesturing, or pitch of voice.
3. Use direction of head to address a service.
4. Mix real sounds with computer generated sounds so user can hear what is happening in the real world.
5. Headphones that allow outside audio in, but can still hear virtual sounds clearly.
6. Bone speakers.
7. Headphones 2" from ear.
8. Bluetooth surgical implants for microphone and speaker.
9. Electronic palate to sense tongue movement.
10. Use acoustics for a size of room that best matches the number of objects in the audio field. If a search result had many matches then would house them in a large room.
11. Use a cathedral sound. Related to 10.
12. Use of reverberation/reflection to simulate room acoustics. Related to 10.
13. Use echo to get idea of position.
14. Process audio to reduce the quality for distant sounds. E.g. limit bandwidth, muffle, distort, add noise, reduce volume. Also used to distinguish virtual from real sounds.
15. Shake removal for direction sensor – apply a hysteresis or dead zone. Idea is to minimize annoyance of objects moving in audio field when not intended.
16. Link movement of objects to head movement.
17. Everything in virtual world deliberately sounds synthetic so users can distinguish the real from the virtual world.

22

18. Make the virtual world fixed relative to the head. (Opposite of 16). This allows users head to move (e.g. to nod or shake head) without the audio space changing. Also helps user distinguish real from virtual.

19. Adjust volume of virtual sounds to match background ambient noise level. E.g. in noisy car, stadium.

20. Audio refresh. User can refresh a part or whole of audio field to hear some audio output again.

21. Audio fade. Repeating sounds or voices that attempt to attract the users attention attenuate away after a few seconds – so not annoying to user. 20 is useful here.

22. Direction can be used to encode importance. E.g. importance goes clockwise.

23. Nod detector. Sense movement to select a sound from a direction. Similar to 3.

24. Rotate audio environment. Use can spin objects to bring sounds to the front, e.g. to select them, or to get better spatial resolution.

25. TDM for sounds. Instead of laying out sounds in space, order them in time. Mono is fine for this.

26. Logical slot expands to physical representation. The 3d position of sounds encodes logical position. One segment of this logical space is dedicated to sounds from the virtual world, such as beacons. User selects this logical segment when hears something interesting to replace and fill the whole audio field with sounds to match objects in physical world.

27. User can mute segments of audio space.

28. Layers in a cylinder. Objects are placed left-right around a circle on the horizontal. There are many circles, brought from layers up or down an imaginary cylinder, that each represent a different context.

29. Position as a mnemonic. For search results could lay out in 3d, to make finding and remembering easier.

30. Lists of items, e.g from a DialogML script, are spaced clockwise, for reasons in 29.

31. Encode hierarchies up and down. E.g individual shops in cribbs causeway are at one level, the cribbs causeway service itself is at a higher level.

23

32. Bookmarks for location in space and set of objects found in that vicinity. Examples: lunch ideas, restaurants. Could return to bookmark and find a room laid out in 3d virtual space that corresponded to the physical arrangement at time of snapshot.

33. Ad hoc 3d speakers. E.g. discover, connect to, and use 4 speaker system in car to position sounds.

34. Content of media stream is independent of 3d positioning.

35. Information pump refill. Group of objects in refill point could be bookmarked or moved around as a unit.

36. (Investigate bandwidth, coding and architecture for effective 3D positioning. E.g. where to do 3D mixing, effect of codecs on frequencies used for positioning.)

37. Poor quality (low-bandwidth) audio can be positioned accurately with high-quality local audio processing.

38. Vary audio stream B/W depending on audio focus – audio streams in front of field get more B/W. Intention is allow possibility to do mixing locally for small number of sounds.

39. Send samples from network voice browser, cache locally, do mixing locally.

40. Send high quality stereo mix of all sounds in audio field, but with less bandwidth than required to do accurate 3D positioning. In addition send small number of low-bandwidth sound channels (with associated 3d position coordinates), so they can be mixed locally with high positioning accuracy.

41. Do most mixing in network voice browser, but send separate low bit rate encoding of the frequencies to do 3D positioning of sounds, that are mixed in at local device.

42. Some sounds need to be positioned in center of field or will be irritating to user – example was certain instruments, such as drums.

43. Audio scanning. Play sounds or music that represent (prepares the user for?) the complexity of information.

44. Handset – finger in ear – hand ear piece using sounds transmitted through bone.

45. Size of room depends on number of voices (same as 10?)

24

46. Adjust acoustics of virtual room to better fit those of the physical room. For example a beacon could transmit a compact parameterized description of the acoustics of the room that contained it. This would allow the virtual objects to better blend into the real sounds, and would aid in accurate positioning.

47. Use a 3D audio cursor that the user can move around to select objects in 3D space.

48. Sounds get louder or change some characteristic when the 3D cursor is near. For example if touch a service it goes 'ooohh'.

49. Have a unique, unchangeable service name that is used to refer to an object. A service or a sound could announce its service name when it fades after the attention timeout.

50. Cursor could hum.

51. Rhythm labels. Use rhythm as a means of distinguishing multiple audio streams or representing meaning. MacDonald's could repeat their name in a pattern 1-23, 1-23.

Physical match of 3D audio

52. Use a sound or speaking voice as an audio navigation aid to find a particular place – sound moves around as I move, know when I am going in right direction or getting closer.

53. User sets a radius for the audio radar. Does this with either visual slider, or spoken commands

54. User sets filter criteria for audio radar. This can be done using either a visual tool (perhaps specifying key words or semantic concepts), or using voice.

55. Services in audio field are discovered using both physical and virtual beacons, and by location-indexed database searches.

56. User can use direction information to narrow down location searches.

57. User can choose to have objects dynamically move in the audio field as the user moves around; the update might be continuous or periodic (e.g. every 30

25

seconds). Alternatively, user can freeze the position of objects in the audio field; the positions will remain frozen until the user explicitly does a resynchronization. This prevents the disorientation caused by unintended movements of the direction sensing device.

58. Objects could be positioned relative to the users current position, but not their current orientation – objects to the North are in front, to the south behind. This way the objects move less wildly; it is up to the user to figure out the direction relative to the orientation.

Logical positioning

59. Use the 3D position of sounds as a way of presenting multiple simultaneous audio streams. Brain is better at disentangling voices that are separated in space.

60. Use position to represent work areas. E.g. behind is unimportant, left-rear is e-mail, front-below is telephone calls, front-up is interruptions such as reminders, etc.

61. Use audio rooms as a way of navigating audio spaces. Little bit like multiple workspaces in UNIX desktop. Each workspace has 3d positioning.

62. Sounds can appear to leak from room to room if sufficiently important. The doors have well known 3D positions.

63. Can use distance and loudness to represent relevance or importance.

64. Use 3D positioning to fill a space with search results from a Web search. The relevance of the search is ordered in a clockwise direction, the time of the response is positioned up or down.

65. Sounds can be processed to mean different things. E.g. sounds that leak from a room might echo, continuous background information might whisper, important information might have a lower pitch.

66. Use different voices for different applications.

67. Use clockwise rotation to time order the services that I start. When I speak to a service it moves to the front position, when I switch to another service it returns to its original position.

26

68. Use sounds to represent that a service wants to speak to user. E.g. if a service without the focus performs a TTS, the request is queued up, but a distinctive repeating sound is generated from the appropriate position for the service, perhaps with the name of the service being whispered.

69. Users can manually change the position of sounds and services.

Audio Interface

70. Services can stream audio to a user, play distinctive repeating sounds, push TTS content with specific voice characteristics. All of the above can be processed, or positioned in 3D.

71. Particularly urgent interruptions or reminders could swirl around the users head until they respond.

Sensing

72. Direction could be sensed a) directly by an electronic compass in a headset, glove, PDA, wrist patch b) calculated from the position history of the user (esp. when in car, e.g. going down M4) c) deduced from the fact the user is looking directly at something, such as a beacon next to a painting.

73. Sensing may not produce continuous updates. 3d positions could remain relative to the last discrete direction reading. E.g. for the beacon example in a museum the 3d positions of other exhibits remains fixed until the user next looks at a beacon – it is up to the user to figure out the discrepancy.

Technology

74. Assume that the VB is physically distinct from the output device such as a PDA. VB is in the network, user carries PDA. VB can do mixing and 3d spatialization

27

of audio and TTS, and pass the result to PDA. The PDA can play this directly, or could optionally locally mix additional audio sources; e.g. local calendar, or info received over local wireless networks.

THIS PAGE BLANK (USPTO)